IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs

Chi-Tung Cheng[1,2] · Tsung-Ying Ho[3] · Tao-Yi Lee[4] · Chih-Chen Chang[5] · Ching-Cheng Chou[1] · Chih-Chi Chen[6] · I-Fang Chung[2,7,8] · Chien-Hung Liao[1,9]

Abstract

Objective To identify the feasibility of using a deep convolutional neural network (DCNN) for the detection and localization of hip fractures on plain frontal pelvic radiographs (PXRs).

Summary of background data Hip fracture is a leading worldwide health problem for the elderly. A missed diagnosis of hip fracture on radiography leads to a dismal prognosis. The application of a DCNN to PXRs can potentially improve the accuracy and efficiency of hip fracture diagnosis.

Methods A DCNN was pretrained using 25,505 limb radiographs between January 2012 and December 2017. It was retrained using 3605 PXRs between August 2008 and December 2016. The accuracy, sensitivity, false-negative rate, and area under the receiver operating characteristic curve (AUC) were evaluated on 100 independent PXRs acquired during 2017. The authors also used the visualization algorithm gradient-weighted class activation mapping (Grad-CAM) to confirm the validity of the model.

Results The algorithm achieved an accuracy of 91%, a sensitivity of 98%, a false-negative rate of 2%, and an AUC of 0.98 for identifying hip fractures. The visualization algorithm showed an accuracy of 95.9% for lesion identification.

Conclusions A DCNN not only detected hip fractures on PXRs with a low false-negative rate but also had high accuracy for localizing fracture lesions. The DCNN might be an efficient and economical model to help clinicians make a diagnosis without interrupting the current clinical pathway.

Key Points
• *Automated detection of hip fractures on frontal pelvic radiographs may facilitate emergent screening and evaluation efforts for primary physicians.*
• *Good visualization of the fracture site by Grad-CAM enables the rapid integration of this tool into the current medical system.*
• *The feasibility and efficiency of utilizing a deep neural network have been confirmed for the screening of hip fractures.*

Keywords Hip fractures · Neural network (computer) · Machine learning · Algorithms

---

I-Fang Chung and Chien-Hung Liao contributed equally to this work.

✉ Chien-Hung Liao
  surgymet@gmail.com

1 Department of Trauma and Emergency Surgery, Chang Gung Memorial Hospital, Linkou, Chang Gung University, Taoyuan, Taiwan

2 Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan

3 Departments of Nuclear Medicine and Molecular Imaging Center, Chang Gung Memorial Hospital, Linkou, Chang Gung University, Taoyuan, Taiwan

4 Donald Bren School of Information and Computer Sciences, University of California Irvine, Irvine, CA, USA

5 Department of Medical Imaging and Intervention, Chang Gung Memorial Hospital, Linkou, Chang Gung University, Taoyuan, Taiwan

6 Departments of Rehabilitation and physical medicine, Chang Gung Memorial Hospital, Linkou, Chang Gung University, Taoyuan, Taiwan

7 Center for Systems and Synthetic Biology, National Yang-Ming University, Taipei, Taiwan

8 Preventive Medicine Research Center, National Yang-Ming University, Taipei, Taiwan

9 Center for Artificial Intelligence in Medicine, Chang Gung Memorial Hospital, Linkou, Taoyuan, Taiwan

Abbreviations

AUC	Area under the curve
CAM	Class activation mapping
CGMH	Chang Gung Memorial Hospital
CGTRP	Chang Gung Trauma Registry Program
CT	Computed tomography
DCNN	Deep convolutional neural network
Grad-CAM	Gradient-weighted class activation mapping
MRI	Magnetic resonance imaging
PXR	Frontal pelvic radiograph
ROC	Receiver operating characteristic

Introduction

Hip fractures and the resulting postsurgical outcomes are significant public health concerns worldwide [1–5]. Hip fracture is a predominant injury in elderly patients, and the incidence is approximately 250,000 per year in the USA and is expected double in 30 years [1, 3, 6, 7]. Although the incidence has declined recently [3, 8], the number of hip fractures will increase due to the prolonged human life span and growing elderly population. As many as 20–30% of people with a hip fracture will die in the subsequent year [5, 9–11], and many will experience significant functional loss [5, 12]. Early diagnosis and management might preserve not only the function of the joint but also the ambulation and quality of life of the patient [13]. Frontal pelvic radiographs (PXRs) are an essential and widely used tool for image evaluation for hip fractures. However, the sensitivity of PXRs for assessing hip fracture is not optimal. Some previous studies showed that initial misdiagnosis was as high as 7–14% [14, 15], and delayed diagnosis and treatment worsen the prognosis [16]. To avoid further health sequelae and medical costs associated with a delayed diagnosis, additional radiographs, nuclear medicine bone scans, computed tomography (CT) scans, and magnetic resonance imaging (MRI) scans have been recommended as routine diagnostics [14, 17]. However, it is not an effective, efficient, or economical method to use these diagnostic tools in routine examinations.

Digital medical imaging systems offer not only immediate and remote access [18] but also the possibility for computer-aided diagnostic procedures [19]. Computerized analysis based on deep learning has shown potential benefits as a diagnostic strategy and has recently become feasible [20]. The application and achievement of deep convolutional neural networks (DCNNs) in the medical field are expected to grow rapidly, and some studies have presented a great opportunity to apply deep learning in trauma [21–24]. DCNN has a proven ability to classify the bone structure of body parts and identify fractures in specific sites with expert-level accuracy [25–28]. Meanwhile, the efficiency and feasibility of using a DCNN for detecting hip fractures have not been evaluated completely.

Although one recent article achieved high accuracy in identifying hip fractures using the region base method [29], the "black box" mechanism of deep learning is still the major hindrance for its clinical application. Many methods for the visual explanation of a DCNN have been developed, such as saliency mapping, class activation mapping (CAM), and gradient-weighted class activation mapping (Grad-CAM) [30]. Visualization mapping of images may validate the lesion detection ability of deep learning algorithms. Grad-CAM with the Keras-vis library generates a heatmap that visualizes the class-discriminative regions. In a medical image, Grad-CAM can help the physician to identify the pathologic region and validate the DCNN performance. Automated detection of hip fractures has potential benefits, such as increasing efficiency, decreasing misdiagnosis, reducing delayed management, and improving patient outcomes, especially in emergent situations if the algorithm achieved expert-level accuracy.

In this study, we developed an automated fracture diagnosis algorithm trained based on the DCNN to examine PXRs and investigated the performance compared with that of the physicians. We also investigated the validity of this algorithm by lesion visualization using Grad-CAM.

Materials and methods

Study population

We utilized the Chang Gung Trauma Registry Program (CGTRP) in Chang Gung Memorial Hospital (CGMH), Linkou, Taiwan. Demographic data, medical data, perioperative procedures, hospital procedures, medical imaging findings, follow-up data, and information regarding complications were recorded prospectively in a computerized database. We extracted the data and images of all trauma patients treated from August 2008 to December 2017 at CGMH, which is a level I trauma center. The Internal Review Board of CGMH approved the study.

Training PXR dataset (2012–2016 PXR dataset)

Patients in the trauma registry treated from January 2012 to December 2016 were evaluated and selected if admitted and had frontal PXRs performed on the date of injury. The PXRs were stored automatically with a Python script for a picture archiving and communication system (PACS) viewer. The size of stored images varies from 2128×2248 pixels to 2688×2688 pixels, and the color is 8-bit grayscale. Each was given a serial number and de-identified in both the images and registry. After the images were stored, all images were evaluated. All labels in each image were carefully examined and removed.

Limb radiograph dataset

The limb radiograph dataset, which included views of the ankles, feet, knees, wrists, and elbows, was collected and de-identified in a similar way as previously described. Each image was labeled as the corresponding body part during the automatic storage process. This dataset was used mainly to develop a pretrained medical imaging model for further transfer learning.

Independent testing PXR dataset (2017 PXR dataset)

Another hundred patients were identified from the trauma registry during 2017 who had PXRs performed. There were 25 patients with femoral neck fractures, 25 with intertrochanteric fractures, and 50 without hip fractures. The PXRs performed on the date of injury were prepared as an independent testing dataset.

Image labeling and preprocessing

After the 2008–2016 PXR datasets were established, the images were initially labeled as a hip fracture or no hip fracture according to the diagnosis in the trauma registry. The radiologist's report, diagnosis, clinical course, and other related images, such as CT or other views of the hip joint, were reviewed if the label was questionable. Poor-quality images such as those with poor image contrast, positioning errors, and foreign body interference and those with other fractures, including femoral shaft fractures and pelvic fractures, were excluded. Each image was reviewed by a trauma surgeon for the preciseness of the label and quality of the images. The limb datasets were labeled automatically as previously described.

Development of the algorithm

DCNNs are widely used in computer vision and medical image recognition [31]. A DCNN is a machine learning algorithm developed from an artificial neural network. The basic concept is to use pixel values from a digital image as inputs using techniques, such as convolution and pooling, on each layer and to adjust the weights in the neural network according to the difference between the output and true label. After a significant amount of imaging input is used as the training material, the weights in the neural network are adjusted to fit the problem. We used DenseNet-121 as the structure of our neural network [32]. The structure contains a dense block with a skip connection designed within the dense block. The input images were resized to 512×512 pixels with an 8-bit grayscale color to reduce the complexity and computation. Most studies use ImageNet as a pretraining material for "transfer learning" [26, 28, 33, 34]. Instead, we used the limb dataset as our pretraining material. The model

was initially trained to identify the body part in view on each limb radiograph. We randomly chose 90% of the limb dataset for training and 10% for validation. The pretrained weights of the DCNN were preserved for PXR training. The PXR dataset was separated as 80% for training and 20% for validation. During the training process, image augmentation was applied with a zoom of 10%, horizontal flip, vertical flip, and rotation of 10°. The batch size was 8, and the Adam optimizer was used. The initial learning rate was 10^{-3} with a reduced learning rate on the plateau. The final model was trained with 60 epochs under the above hyperparameters.

Evaluating the algorithm

The trained hip model was tested with the independent 2017 PXR dataset to evaluate its accuracy for identifying hip fractures. The probability generated by the model of hip fracture was evaluated with a receiver operating characteristic (ROC) curve and the area under the curve (AUC). A confusion matrix was also calculated using a cutoff level of probability 0.5 of hip fracture. For those PXR models that predict fractures, we also used a Grad-CAM to generate a heatmap that the model activated for the hip fracture to provide evidence that the model indeed recognized the fracture site. The heatmaps were also reviewed by a radiologist to compare with the fracture site on the original image to evaluate the ability of localization. We also recruited experts from the surgical, orthopedics, emergency, and radiology departments to evaluate the accuracy of each subspecialist. A web-based questionnaire for the 2017 PXR dataset was designed with a multiple-choice question to answer whether or not there was a hip fracture in the image.

Statistical analysis and software

The software used to build the DCNN was based on an Ubuntu 14.04 operating system with TensorFlow 1.5.1 and Keras 2.1.4 open-source library with Python 3.6.5 (Python Software Foundation). The training process was run on an Intel® Core™ i7-7740X CPU 4.30 GHz with GeForce® GTX 1080 Ti GPU. All statistical analyses were performed using R 3.4.3 with extension packages "pROC," "epiR," "Publish," and "ggplot2." Continuous variables were evaluated with Student's t test, and categorical variables were evaluated with the chi-square test. We compared the hip model and specialists using the sensitivity, specificity, false-negative rate, and false-positive rate, and the F1 scores and 95% confidence intervals (CIs) were calculated. ROC curves and AUCs were used to evaluate the performance of the model.

Results

The demographic data of the patients in the 2008–2016 PXR dataset are shown in Table 1. We obtained 3605 PXRs to build the model and collected 25,505 different limb radiograph views from 12,583 patients, including radiographs of 6019 ankles, 3832 elbows, 4134 feet, and 3378 wrists.

The model was trained with a limb radiograph dataset and finally showed an accuracy of 99.5% on the validation dataset for identifying specific body parts. The weights in the limb model were preserved and named the "limb pretrained model." The model was retrained again using the weights from the limb pretrained model, and the task was changed to detect hip fractures. The final training accuracy values for the training set was 0.94 and for the validation set was 0.90, respectively. The change of accuracy and loss during the training process are shown in Fig. 1.

After applying the hip model to the 2017 PXR dataset, the accuracy, sensitivity, specificity, false-negative rate, and F1 score of the model were 91% ($n = 100$; 95% CI, 84–96%), 98% (95% CI, 89–100%), 84% (95% CI, 71–93%), 2% (95% CI, 0.3–17%), and 0.916 (95% CI, 0.845–0.956), respectively. A total of 21 experts completed the questionnaire. The range of sensitivity of primary physicians (except radiologists and orthopedic surgeons) was 84% to 100% (mean, 95.6%; 95% CI, 93.6–97.6%), and the specificity ranged from 46 to 94% (mean, 82.2%; 95% CI, 76.2–88.3%). The experts, including two radiologists and four orthopedic surgeons, completed the questionnaire and achieved a mean sensitivity of 99.3% (95% CI, 98.2–100%) and a specificity of 87.7% (95% CI, 81.5–93.8%). The ROC curve of prediction probability compared with individual specialist performance is shown in Fig. 2. The model achieved an AUC of 0.98 (95% CI, 0.96–1.00). The 95% CI of both mean sensitivity and mean specificity of the primary physicians were below the ROC curve of the model, and the mean performance of radiologists and orthopedic surgeons was still slightly better than that of the model.

After analyzing 49 heatmap images, the model predicted positive hip fracture. Only two images identified the wrong activation site, and 95.9% of the activation area was located at the hip fracture site, as shown in Fig. 3.

Discussion

These results indicate that a DCNN can be trained to identify hip fractures within image datasets with high sensitivity (98%) and accuracy (91%) (F1 score, 0.91). Currently, hip fractures occur daily, and uneven or nondisplaced hip fractures are challenging to rapidly identify because of the limitations of the human eye and PXRs [14, 15]. With the assistance of the DCNN, we can detect hip fractures immediately with a low false-negative rate (0.02), which is noninferior to the performance of the experts. This DCNN will be useful for primary physicians to lessen the misdiagnosis rate and to prevent subsequent misdiagnosis events. Hip fractures are a promising target for deep learning approaches because of the availability of near-perfect ground truth labels. Because of the weight-bearing nature of the region, patients who have clinically "silent" fractures rapidly develop severe pain and immobility. Early detection and surgery are critical for patient survival and the preservation of hip function. Postponed management of hip fractures results in a poor prognosis and even an increased risk of death years later [35–38]. Therefore, detecting hip fractures as soon as possible is critical for remote mortality and medical outcomes.

DCNNs allow computers to learn from iterations with automatic feature extraction under limited programming, and the prediction rate is highly accurate. Artificial intelligence and the automation of bony fracture detection have been discussed [25, 26, 28]; Gale et al also described a DCNN method that was used for predicting hip fractures [29]. However, the model developed previously for hip fractures is a region base model and requires a localization network to identify the femoral neck first. Our study shows that the detection and diagnosis of hip fractures on PXRs could be performed by the input of a whole-scale radiograph to a DCNN without identifying the specific region first. The deep learning algorithm also achieved an accuracy level that is compatible with the accuracy of radiologists and orthopedic surgeons. Moreover, this study also substantially added to other current studies.

First, we applied a transferring learning method to develop our algorithm. We set the pretrained model using 25,505 unlabeled limb radiographs instead of the Image-Net images

Table 1 The demographic data of the frontal pelvic radiograph dataset

	With hip fracture	Without hip fracture	p value
Number of patients	1975	1630	
Age, mean, years (SD)	72.34 (16.73)	44.88 (20.46)	< 0.001
Gender (% male)	829 (42.0)	1112 (68.2)	< 0.001
ISS, mean (SD)	9.96 (4.21)	14.01 (9.29)	< 0.001
Type of fracture			
Femoral neck fracture	931 (47.1)	NA	
Trochanteric fracture	1044 (52.9)	NA	

SD standard deviation, *ISS* injury severity score
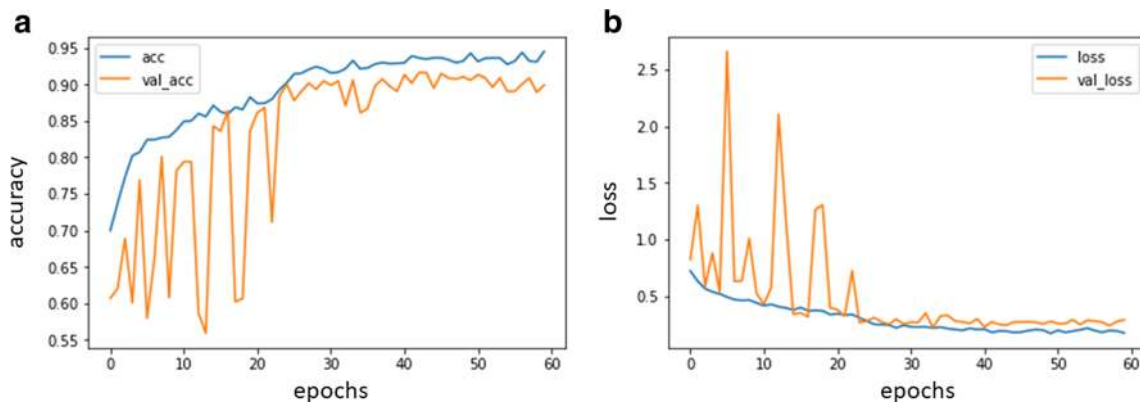
Fig. 1 Performance in the training and validation datasets using TensorBoard. a. Accuracy change during training process. acc, accuracy of the training set; val_acc, accuracy of the validation set;b. Change of loss during training process. loss, loss of the training set; val_loss, loss of the validation set

because we believe that a pretrained model using similar images reduces the required image sample size and training time. This study does not compare performance between the Image-Net pretrained model and the limb pretrained model. We entered 2804 frontal PXRs as training material. After evaluation, our accuracy increased from 79% (scratch pretrained) to 91% (limb dataset pretrained), as we expected, and the pretraining material also impacted the final accuracy.

Unlike in previous works, the process did not require much extensive processing, lesion segmentation, or extraction of domain-specific visual features. In contrast, our system needs no handcrafted features, and it is trained end to end directly from image labels and raw pixels. Our fully automated system takes PXRs and automatically detects the presence of hip fractures. These results demonstrate that deep neural networks can

be trained using sizable non-pixel labeled datasets without having to specify lesion-based features. Our research shows that despite the challenges specific to radiographic data, the development of large, clean datasets is sufficient to achieve high-level automated performance with deep learning systems. In this way, we can save the time associated with segmentation and labeling.

One paradox in DCNNs for analyzing medical images is the "black box" mechanism. The model may use another part of the image rather than the true lesion site to produce the answer. Therefore, visualization of the features became a solution to realizing the underlying mechanism of DCNNs [39]. In this study, we used whole PXRs for training and testing. Then, we performed Grad-CAM to visualize the class-discriminative regions as the fracture sites that DCNN

Fig. 2 Performances of the hip model and physicians. The blue, green, yellow, and red spots indicate the performance of radiologists, surgeons, orthopedic doctors, and emergency physicians, respectively
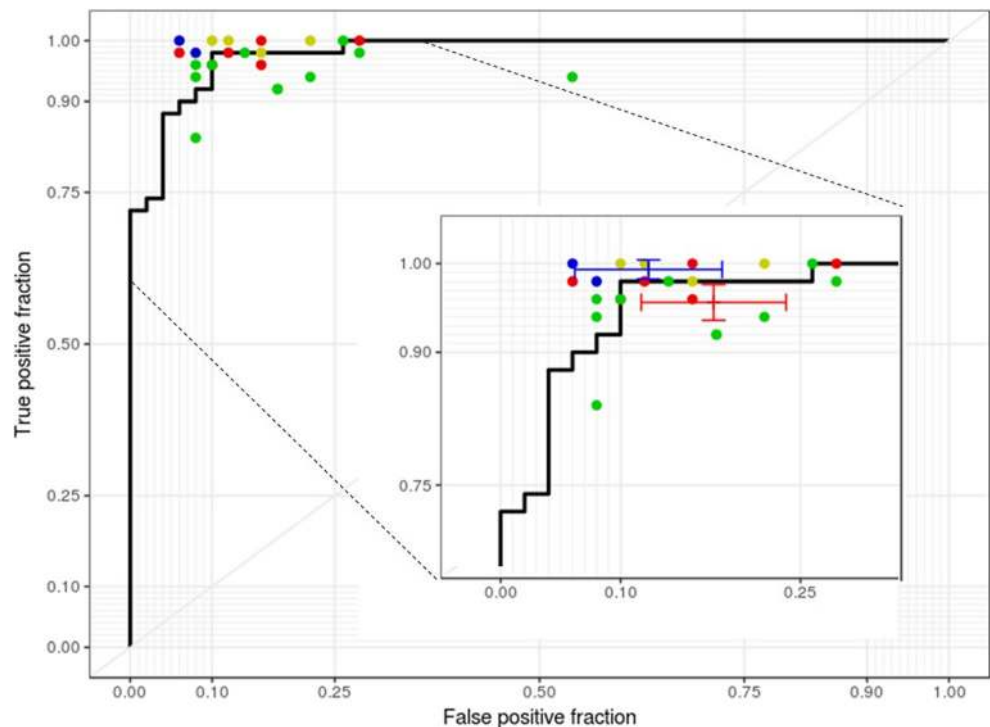
Fig. 3 Grad-CAM-assisted image identification of hip fractures. **a** The original pelvic radiograph with a mildly displaced right femoral neck fracture (arrow) and **b** the image generated after applying the model with Grad-CAM, which visualizes the class-discriminative regions, as the fracture site. **c** PXR presenting a right total hip replacement with a left femoral neck fracture (arrow) and **d** a Grad-CAM-assisted image. **e** PXR with a mildly displaced left femoral neck fracture and **f** a Grad-CAM-assisted image
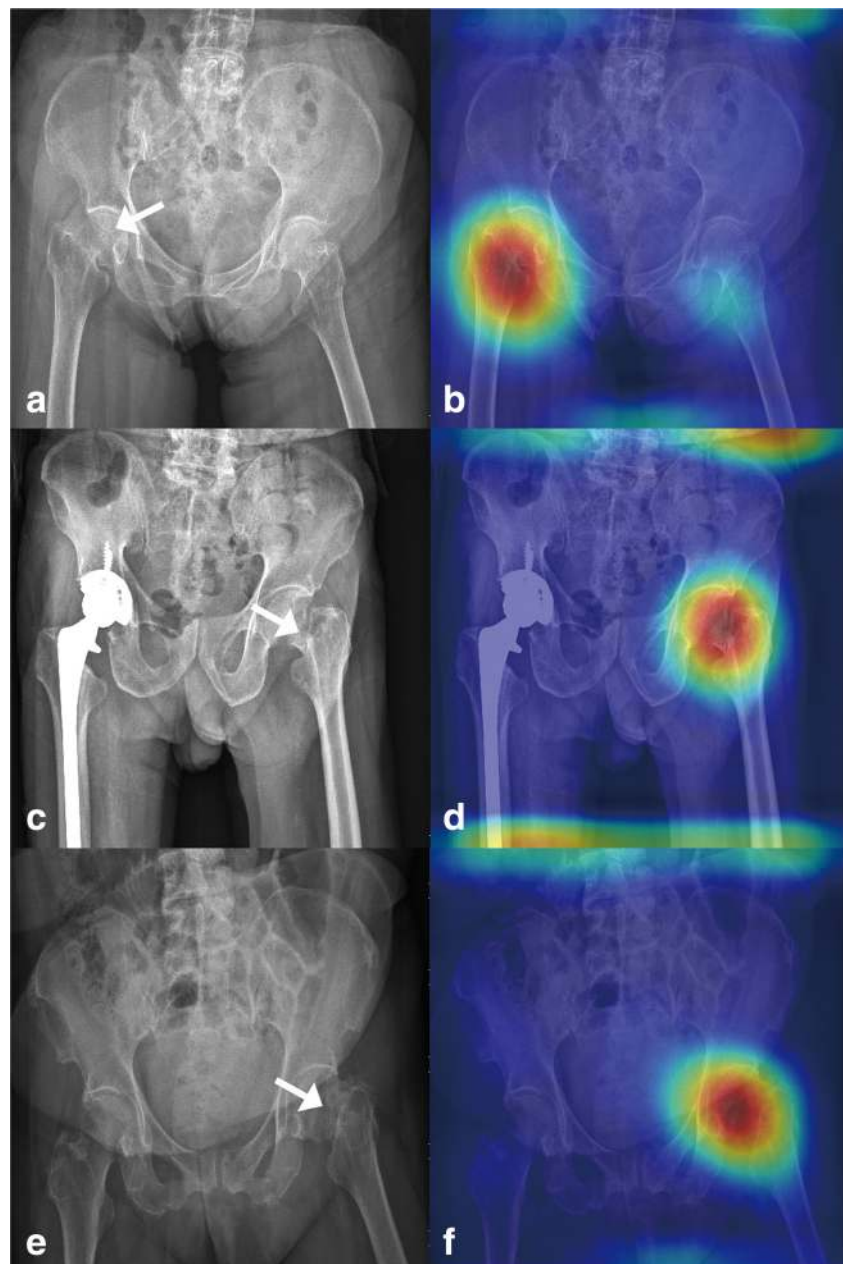

recognizes in the PXRs (Fig. 3). On the other hand, in normal PXRs, the Grad-CAM films lacked a heatmap (Fig. 4). After visualization by the Grad-CAM method, 95.9% of the class-discriminative regions contained the fracture site, which provided evidence that the model indeed recognized the hip fracture. To understand what a DCNN uses to make predictions is a very active research topic in medical aspects, and it may convince doctors to accept the results because the DCNN can explain what they find.

Most deep learning works evaluating medical images use cropped images to avoid "black box" mechanisms and enhance the accuracy of final validation [29, 32].

Once the target is cropped to include the necessary features for recognition, the DCNN will detect the lesion more easily and quickly. In this study, instead of cropping images, we reduced the image matrix size to 512 × 512 pixels. We prefer to input whole PXR images because this method might be more instinctual and physicians are more familiar with it. Because we integrated the DCNN into the clinical pathway, doctors will prefer to use whole images rather than cropped images. The dimensionality reduction also decreases the computational requirement and shortens the training time with an acceptable result. Furthermore, based on our model, we also applied this algorithm to other types

Fig. 4 Grad-CAM-assisted image of a normal PXR. **a** Frontal PXR without a hip fracture and **b** a Grad-CAM image with no heatmap. **c** Frontal PXR with a right hip replacement without a hip fracture and **d** a Grad-CAM image showing no identification of a fracture site. **e** Frontal PXR without a hip fracture and **f** Grad-CAM visualized no class-discriminative regions
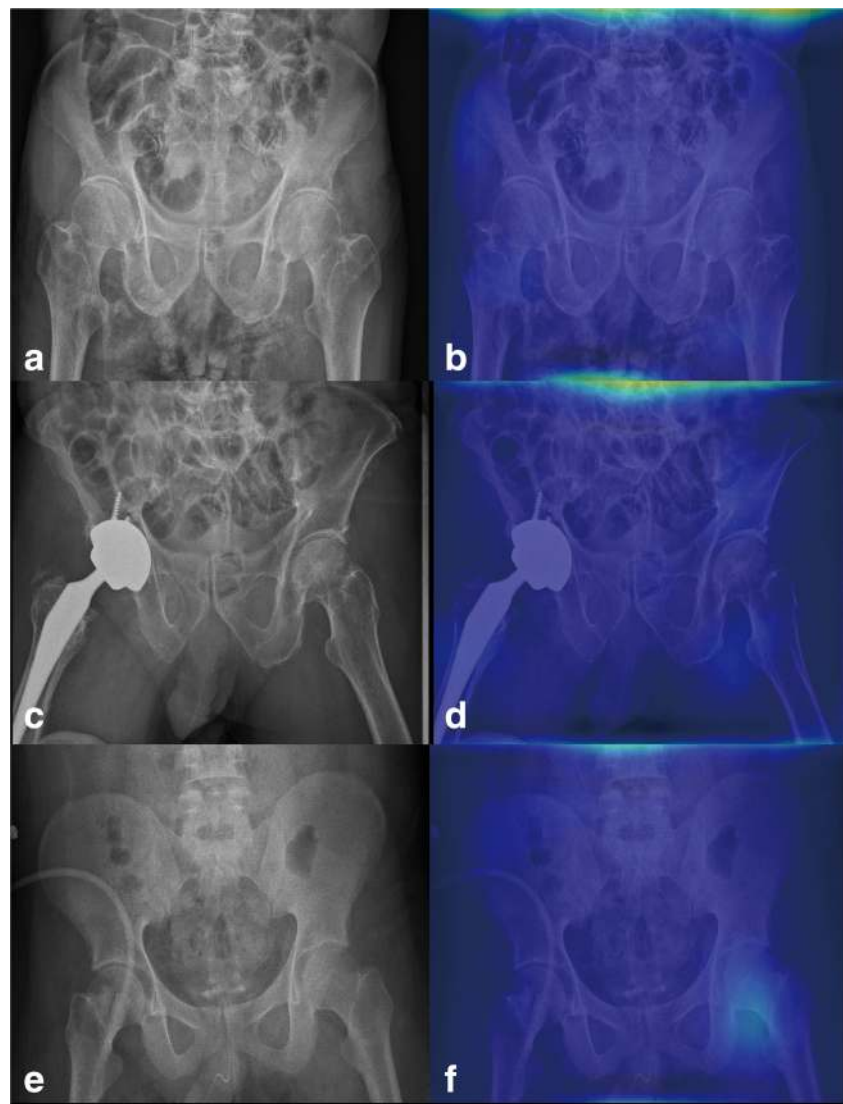
of fracture in PXRs. In the future, based on the observations from this study, the development of similar high-performance algorithms for medical imaging might shorten the training process time and resources required.

This study has several limitations. One fundamental limitation arises from the nature of DCNNs because the neural network was provided with only an image and associated diagnosis, without explicit definitions of features. Because the DCNN "learned" the features that were most predictive, the algorithm might use features previously unknown to or ignored by humans. Although this study showed good visualization for identifying fractures, the exact features being used are still unknown. It is possible that the heatmaps show differences in femoral alignment or soft tissue contrast due to edema or hemorrhage between the fractured and nonfractured sites. In this study, two of the images did not demonstrate the right activation location of the fracture. One of the wrong activation sites is located on the opposite hip, and another

had a stronger signal over the iliac bone region. This error is indeed a limitation of DCNN because it is difficult to explain why it activates at the wrong site. Inputting the whole image to the model is also the most challenging part, and we will exert further effort to solve this issue using a greater number of input samples. In addition, the algorithm was specifically trained to discriminate between healthy bones and fractures in the background of the bony architecture on radiographs, but the algorithm might be unable to identify other pathological presentations. The detection of other lesions on PXRs, which is relevant for routine diagnoses, was not included in this study. In the current study, although we used only PXR as the input material, we still observed differences in age, gender, and ISS between both fracture and nonfracture patients, which might create some selective bias. Furthermore, we extracted the hip fracture images by final diagnosis from the registry and identified the fracture site on the images when we validated the performance of the

DCNN. Therefore, it is difficult to clarify the exact occult fracture rate. In this study, we still manually excluded a certain proportion of images, which will be problematic when trying to incorporate this process into the clinical workflow. Finally, although the results of this study are promising, integrating this automatic detection algorithm into clinical work to increase the detection rate of hip fractures presents another challenge. Therefore, we utilized a web-based system that can input the PXR from PACS in the hospital, and the DCNN will detect the presence of a fracture or not and localize the fracture region by Grad-CAM. A randomized, prospective study should be conducted to evaluate the clinical impact on the diagnostic accuracy and economic value of DCNN for identifying hip fractures.

In conclusion, to identify hip fractures on PXRs, the algorithm trained by a DCNN achieved excellent performance with a high accuracy and low false-negative rate and is useful as a screening tool. Furthermore, our algorithm can localize the fracture site with high accuracy and thus assist clinical physicians in identifying more occult hip fractures and managing patients with these fractures early to prevent further medical costs and decreased quality of life.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Dr. Chi-Hsun Hsieh.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry Dr. Chi-Chih Chen, who has significant statistical expertise, kindly provided statistical advice for this manuscript.
No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained as CGMHIRB 201702059B0.

Study subjects or cohort overlap Some study subjects or cohorts have not been previously reported.

Methodology
• Retrospective

• Diagnostic or prognostic study
• Performed at one institution

References

1. Brauer CA, Coca-Perraillon M, Cutler DM, Rosen AB (2009) Incidence and mortality of hip fractures in the United States. JAMA 302:1573–1579. https://doi.org/10.1001/jama.2009.1462
2. Riggs BL, Melton LJ III (1995) The worldwide problem of osteoporosis: insights afforded by epidemiology. Bone 17:S505–S511. https://doi.org/10.1016/8756-3282(95)00258-4
3. Leslie WD, O'Donnell S, Jean S et al (2009) Trends in hip fracture rates in Canada. JAMA 302:883–889. https://doi.org/10.1001/jama.2009.1231
4. Bliuc D, Nguyen ND, Milch VE et al (2009) Mortality risk associated with low-trauma osteoporotic fracture and subsequent fracture in men and women. JAMA 301:513–521. https://doi.org/10.1001/jama.2009.50
5. Center JR, Nguyen TV, Schneider D et al (1999) Mortality after all major types of osteoporotic fracture in men and women: an observational study. Lancet 353:878–882. https://doi.org/10.1016/S0140-6736(98)09075-8
6. Auron-Gomez M, Michota F (2008) Medical management of hip fracture. Clin Geriatr Med 24:701–19– ix. https://doi.org/10.1016/j.cger.2008.07.002
7. Marks R (2010) Hip fracture epidemiological trends, outcomes, and risk factors, 1970-2009. Int J Gen Med 3:1–17
8. Michael Lewiecki E, Wright NC, Curtis JR et al (2018) Hip fracture trends in the United States, 2002 to 2015. Osteoporos Int 29:717–722. https://doi.org/10.1007/s00198-017-4345-0
9. Melton LJ, Gabriel SE, Crowson CS et al (2003) Cost-equivalence of different osteoporotic fractures. Osteoporos Int 14:383–388. https://doi.org/10.1007/s00198-003-1385-4
10. Johnell O, Kanis J (2005) Epidemiology of osteoporotic fractures. Osteoporos Int 16(Suppl 2):S3–S7. https://doi.org/10.1007/s00198-004-1702-6
11. Johnell O, Kanis JA (2006) An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. Osteoporos Int 17:1726–1733. https://doi.org/10.1007/s00198-006-0172-4
12. Boonen S, Autier P, Barette M et al (2004) Functional outcome and quality of life following hip fracture in elderly women: a prospective controlled study. Osteoporos Int 15:87–94. https://doi.org/10.1007/s00198-003-1515-z
13. Grimes JP, Gregory PM, Noveck H et al (2002) The effects of time-to-surgery on mortality and morbidity in patients following hip fracture. Am J Med 112:702–709. https://doi.org/10.1016/S0002-9343(02)01119-1
14. Hakkarinen DK, Banh KV, Hendey GW (2012) Magnetic resonance imaging identifies occult hip fractures missed by 64-slice computed tomography. J Emerg Med 43:303–307. https://doi.org/10.1016/j.jemermed.2012.01.037
15. Chellam WB (2016) Missed subtle fractures on the trauma-meeting digital projector. Injury 47:674–676. https://doi.org/10.1016/j.injury.2015.11.001

16. Tarrant SM, Hardy BM, Byth PL et al (2014) Preventable mortality in geriatric hip fracture inpatients. Bone Joint J 96-B:1178–1184. https://doi.org/10.1302/0301-620X.96B9.32814

17. Rehman H, Clement RGE, Perks F, White TO (2016) Imaging of occult hip fractures: CT or MRI? Injury 47:1297–1301. https://doi.org/10.1016/j.injury.2016.02.020

18. Lauro GR, Cable W, Lesniak A et al (2013) Digital pathology consultations—a new era in digital imaging, challenges and practical applications. J Digit Imaging 26:668–677. https://doi.org/10.1007/s10278-013-9572-0

19. Petrick N, Sahiner B, Armato SG et al (2013) Evaluation of computer-aided detection and diagnosis systems. Med Phys 40:087001. https://doi.org/10.1118/1.4816310

20. Kermany DS, Goldbaum M, Cai W et al (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172:1122–1131.e9. https://doi.org/10.1016/j.cell.2018.02.010

21. Prevedello LM, Erdal BS, Ryu JL et al (2017) Automated critical test findings identification and online notification system using artificial intelligence in imaging. Radiology 285:923–931. https://doi.org/10.1148/radiol.2017162664

22. Stone JR, Wilde EA, Taylor BA et al (2016) Supervised learning technique for the automated identification of white matter hyperintensities in traumatic brain injury. Brain Inj 30:1458–1468. https://doi.org/10.1080/02699052.2016.1222080

23. Arif Al SMMR, Knapp K, Slabaugh G (2018) Fully automatic cervical vertebrae segmentation framework for X-ray images. Comput Methods Prog Biomed 157:95–111. https://doi.org/10.1016/j.cmpb.2018.01.006

24. Ho-Le TP, Center JR, Eisman JA et al (2017) Prediction of hip fracture in post-menopausal women using artificial neural network approach. Conf Proc IEEE Eng Med Biol Soc 2017:4207–4210. https://doi.org/10.1109/EMBC.2017.8037784

25. Chung SW, Han SS, Lee JW et al (2018) Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Orthop 89:468–473. https://doi.org/10.1080/17453674.2018.1453714

26. Urakawa T, Tanaka Y, Goto S et al (2018) Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. Skeletal Radiol 41:1–6. https://doi.org/10.1007/s00256-018-3016-3

27. Olczak J, Fahlberg N, Maki A et al (2017) Artificial intelligence for analyzing orthopedic trauma radiographs. Acta Orthop 88:581–586. https://doi.org/10.1080/17453674.2017.1344459

28. Kim DH, MacKinnon T (2018) Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clin Radiol 73:439–445. https://doi.org/10.1016/j.crad.2017.11.015

29. Gale W, Oakden-Rayner L, Carneiro G, et al (2017) Detecting hip fractures with radiologist-level performance using deep neural networks. arXiv:1711.06504

30. Selvaraju RR, Cogswell M, Das A, et al (2017) Grad-CAM: visual explanations from deep networks via gradient-based localization. ICCV 618–626. arXiv:1610.02391

31. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444. https://doi.org/10.1038/nature14539

32. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2018) Densely connected convolutional networks. arXiv:1608.06993

33. Gulshan V, Peng L, Coram M et al (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316:2402–2410. https://doi.org/10.1001/jama.2016.17216

34. Esteva A, Kuprel B, Novoa RA et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542:115–118. https://doi.org/10.1038/nature21056

35. Bottle A, Aylin P (2006) Mortality associated with delay in operation after hip fracture: observational study. BMJ 332:947–951. https://doi.org/10.1136/bmj.38790.468519.55

36. Vidán MT, Sánchez E, Gracia Y et al (2011) Causes and effects of surgical delay in patients with hip fracture: a cohort study. Ann Intern Med 155:226–233. https://doi.org/10.7326/0003-4819-155-4-201108160-00006

37. Shiga T, Wajima Z, Ohe Y (2008) Is operative delay associated with increased mortality of hip fracture patients? Systematic review, meta-analysis, and meta-regression. Can J Anaesth 55:146–154. https://doi.org/10.1007/BF03016088

38. Gdalevich M, Cohen D, Yosef D, Tauber C (2004) Morbidity and mortality after hip fracture: the impact of operative delay. Arch Orthop Trauma Surg 124:334–340. https://doi.org/10.1007/s00402-004-0662-9

39. Erhan D, Bengio Y, Courville A, et al (2009) Visualizing higher-layer features of a deep network. Available via https://pdfs.semanticscholar.org/65d9/94fb778a8d9e0f632659fb33a082949a50d3.pdf. Accessed on 28 Feb 2019